

Maschinelles Lernen II

5. Übung

Prof. Tobias Scheffer
Dr. Niels Landwehr
Christoph Sawade

Sommer 2010

Ausgabe am: 26.05.10
Besprechung am: 02.06.10

Aufgabe 1

k-Means

Gegeben seien folgende Instanzen $\mathbf{x}_1, \dots, \mathbf{x}_4$.

i	$[\mathbf{x}_i]_1$	$[\mathbf{x}_i]_2$
1	0,98	0,18
2	0,33	0,95
3	0,69	0,72
4	0,88	0,48

1. Wenden Sie den k -means Algorithmus mit $k = 2$ an. Welche Initialisierung benutzen Sie? Welche Cluster entstehen? Veranschaulichen Sie Ihre Lösung in einem Koordinatensystem.
2. Geben Sie eine neue Initialisierung an, mit der der k -means Algorithmus andere Cluster erzeugt.
3. Bestimmen Sie das Optimum der beiden erhaltenen Lösungen.

Aufgabe 2

Global optimales Clustering

In der Vorlesung wurde erklärt, dass der k -means-Algorithmus nur ein lokales Optimum liefert. Wie könnte ein naiver Algorithmus aussehen, der das Minimum der Funktion

$$J = \sum_{n=1}^N \sum_{k=1}^K r_{nk} \|\mathbf{x}_n - \mu_k\|^2$$

findet? Analysieren Sie den Aufwand des Verfahrens. Zeigen Sie, dass es exponentiell viele Aufteilungen von n Beispielen in k (auch leere) Cluster gibt.

Aufgabe 3*Mixture of Gaussians*

Der EM-Algorithmus besteht aus zwei sich wiederholenden Schritten: Die Berechnung der Clusterzugehörigkeiten für die Beispiele (*expectation*) und die darauf folgende Maximierung der Modell-Parametern $\Theta = (\pi, \mu, \Sigma)$ (*maximization*). π_k ist dabei der geschätzte Anteil von Punkten im Cluster k . Zeigen Sie für den binären Fall ($k = 2$), dass

$$\pi_k = \frac{N_k}{N}$$

die Zielfunktion $\mathcal{Q}(\Theta, \Theta_t)$ maximiert.